MACHINE LEARNING
HOMEWORK ASSIGNMENT 1

PART II
PROGRAMMING AND QUESTIONS

Answer 1.
The estimated value of P (1)= 0.401801

Answer 2.
The estimated value of P (0)= 0.598199

Answer 3.
The estimated values for mean and variance for the Gaussian corresponding
to attribute capital run length longest and class 1 (Spam) are
mean = 65.293757
variance = 37389.428067

Answer 4.
The estimated values for mean and variance for the Gaussian corresponding
to attribute char_freq_; and Class 0 are
mean = 0.048426
variance = 0.088306

Answer 5.
The prediction for the first 5 examples in the test set are
FOR TEST EXAMPLE 1 CLASS IS 0
FOR TEST EXAMPLE 2 CLASS IS 0
FOR TEST EXAMPLE 3 CLASS IS 0
FOR TEST EXAMPLE 4 CLASS IS 0
FOR TEST EXAMPLE 5 CLASS IS 0

Answer 6.
The prediction for the last 5 examples in the test set are
FOR TEST EXAMPLE 196 CLASS IS 0
FOR TEST EXAMPLE 197 CLASS IS 0
FOR TEST EXAMPLE 198 CLASS IS 0
FOR TEST EXAMPLE 199 CLASS IS 0
FOR TEST EXAMPLE 200 CLASS IS 0

Answer 7.
The percentage error on the examples in the test file is 20%

Answer 8.
The accuracy attained by Zero-R is 59% as opposed to 80% provided by Gaussian
Naive Bayes on this training set.

Answer 9.
Yes, these assumptions are reasonable for the
spam dataset as the attributes are independent of each other and have continuous values.